Deep sequencing identifies hepatitis B virus core protein signatures in chronic hepatitis B patients

van der Ree, Meike H.; Jansen, Louis; Welkers, Matthijs R. A.; Reesink, Hendrik W.; Feenstra, K. Anton; Kootstra, Neeltje A.

published in
Antiviral research
2018

DOI (link to publisher)
10.1016/j.antiviral.2018.08.009

document version
Peer reviewed version

Title: Deep sequencing identifies hepatitis B virus core protein signatures in chronic hepatitis B patients

Running title: Hepatitis B virus core protein signatures

Meike H. van der Ree [1,2], Louis Jansen [1,2], Matthijs R.A. Welkers [3], Hendrik W. Reesink[1,2], K. Anton Feenstra[¶ 4], Neeltje A. Kootstra[¶ 2]

[¶] Participated equally in the study

1. Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, The Netherlands
2. Department of Experimental Immunology, Academic Medical Center, Amsterdam, The Netherlands
3. Department of Medical Microbiology, Academic Medical Center, Amsterdam, The Netherlands
4. Center for Integrative Bioinformatics VU (IBIVU), Department of Computer Science, and Amsterdam Institute for Molecules, Medicine and Systems (AIMMS), VU University Amsterdam

Electronic word count

No. of words (introduction-discussion):	4853 words
No. of words (abstract):	151 words

Number of figures and tables

No. of tables:	3
No. of figures:	3
No. of tables appendix:	4
No. of figures appendix:	3

Corresponding author
Neeltje Kootstra, PhD
Meibergdreef 9
1105 AZ Amsterdam
Telephone number: +31 20 566 8298
Fax number: +31 20 566 9701
E-mail: n.a.kootstra@amc.uva.nl

Accepted for publication in Antiviral Research

ABSTRACT

Background: We aimed to identify HBc amino acid differences between subgroups of chronic hepatitis B (CHB) patients.

Methods: Deep sequencing of HBc was performed in samples of 89 CHB patients (42 HBeAg positive, 47 HBeAg negative). Amino acid types were compared using Sequence Harmony to identify subgroup specific sites between HBeAg-positive and -negative patients, and between patients with combined response and non-response to peginterferon/adefovir combination therapy.

Results: We identified 54 positions in HBc where the frequency of appearing amino acids was significantly different between HBeAg-positive and -negative patients. In HBeAg negative patients, 22 positions in HBc were identified which differed between patients with treatment response and those with non-response. The fraction non-consensus sequence on selected positions was significantly higher in HBeAg-negative patients, and was negatively correlated with HBV DNA and HBsAg levels.

Conclusions: Sequence Harmony identified a number of amino acid changes associated with HBeAg-status and response to peginterferon/adefovir combination therapy.

KEYWORDS

Hepatitis B virus core protein, chronic hepatitis B, deep sequencing, HBeAg status, treatment response

INTRODUCTION

Hepatitis B virus (HBV) infection is considered as a major public health issue with an estimated 240 million people worldwide being chronically infected. In chronic hepatitis B (CHB) patients, the immune systems fails to clear the virus and patients are at increased risk for developing liver-related complications such as liver cirrhosis and hepatocellular carcinoma (Dienstag, 2008). Two types of drugs are approved for the treatment of CHB patients: nucleos(t)ide analogues (NUCs) and pegylated interferon (peg-IFN). Although NUCs efficiently block viral replication, treatment rarely leads to a functional cure (defined as hepatitis B surface antigen (HBsAg) loss with or without the formation of anti-HBs antibodies) (Chevaliez et al., 2013). In CHB patients treated with peg-IFN a functional cure is only achieved in 3-7% of patients (Janssen et al., 2005; Marcellin et al., 2004). Improved understanding of the replication cycle of HBV resulted in the development of new antiviral drugs that directly inhibit viral replication (Durantel and Zoulim, 2016; Tong and Revill, 2016).

The HBV core protein (HBc) is a promising antiviral target since it is involved in almost all steps of the HBV replication cycle (Summers and Mason, 1982; Tong and Revill, 2016). HBc consists of an assembly domain (residues 1 to 149), and a C-terminal domain (CTD) (residues 150 to 183 or 185, depending on HBV genotype), which are connected by a linker region (residues 141 to 149) (Birnbaum and Nassal, 1990; Watts et al., 2002; Yu et al., 2013; Zlotnick et al., 1997). HBc self-assembles to form viral capsids, predominantly consisting of 120 HBc homodimers (Zlotnick et al., 1996), wherein pregenomic RNA (pgRNA) is reverse transcribed to relaxed circular DNA (rcDNA). Viral capsids containing rcDNA are either encapsulated by the viral envelope proteins and secreted as virions, or they are shuttled back to the nucleus where the rcDNA is released and converted to covalently closed circular DNA (cccDNA) (Summers and Mason, 1982). The cccDNA is the template for transcription of viral mRNAs utilizing host DNA-dependent RNA polymerase. Next to virus assembly, HBc is involved in regulation of cccDNA and host-gene expression in the nucleus of infected hepatocytes (Guo et al., 2011; Li et al., 2010).

HBV replication through an RNA intermediate causes a high mutation rate during chronic HBV infection (Akarca and Lok, 1995; Chuang et al., 1993). Mutations in the basal core promotor (BCP) and precore (PC) regions are most extensively studied, and abrogate the production of hepatitis B e antigen (HBeAg) (Alexopoulou et al., 1997; Buckwold et al., 1996; Carman et al., 1989; Lok et al., 1994), and were shown to influence peg-IFN based treatment response in CHB patients (Erhardt et al., 2000; Jansen et al., 2017; Marrone et al., 2003; Sonneveld et al., 2012). Previously, it was postulated that mutations in HBc could induce secretion of virions containing an immature viral genome, affect envelopment of viral capsids, and may evade immune recognition (Akarca and Lok, 1995; Hosono et al., 1995; Koschel et al., 2000; Yuan et al., 1999a; Yuan et al., 1999b). However, most previous studies analysing HBc mutants used population based sequencing techniques which limits the detection of mutations with low frequencies. Importantly, data on the effect of minority variants of HBc and their effect on HBeAg status and treatment response are lacking. In our cohort of HBeAg-positive and -negative CHB patients treated with a combination of peg-IFN and adefovir for 48 weeks, a high rate of HBsAg loss was observed (11-17% at year 2 of treatment-free follow-up) (Takkenberg et al., 2013). Here, we studied amino acid differences in HBc in our cohort of CHB patients using deep sequencing. We aimed to identify amino acids differences in HBc between HBeAg positive and negative patients. Furthermore, we aimed to identify HBc signatures associated with treatment response in CHB patients.

MATERIALS AND METHODS

Study population

Of the 92 CHB patients who were treated for 48 weeks with peginterferon alfa-2a 180 µg subcutaneously once a week, and adefovir dipivoxil 10 mg daily in the initial investigator-initiated study (controlled-trials.com; ISRCTN 77073364) (Takkenberg et al., 2013), 89 had baseline plasma samples available for sequencing analysis. 84/89 patients completed 48 weeks of treatment and 2 years of follow-up, and comprised the per-protocol population to study associations with treatment response **(Appendix Figure 1)**. Achievement of combined response (HBeAg negativity, HBV DNA levels ≤ 2,000 IU/mL, persistent normal alanine aminotransferase (ALT) levels) was determined 24 weeks (week 72) of treatment-free follow-up.

Deep sequencing

For each patient a pre-treatment (baseline) plasma sample was subjected to 454 ultra-deep pyrosequencing. HBV DNA was isolated from 200 µl plasma by the MagNA Pure LC instrument, using the Total Nucleic Acid isolation kit (Roche Applied Science). DNA was used in a full HBV genome amplification, in an overlapping amplicon approach, using degenerative PCR fusion primer sets to cover the HBV genome (**Appendix Table 1**). For full genome coverage, 12 separate overlapping amplicons were generated with an average length of 400 nucleotides each (**Appendix Table 1**). The obtained PCR fragments were subjected to deep sequencing using the Roche 454 platform according to the manufacturer's instructions. The entire workflow of PCR and 454 sequencing was performed by 454 Life Sciences, Branford CT. First, primer sequences were removed by trimming the first and last 30 nucleotides of each sequence read followed by removal of low quality nucleotides (phred < 28) from the 3' read end. Subsequently, final read quality was checked using the quality control function from QUASR v.7.0.1 (with setting –m 33 and –l 50). Next, all trimmed and quality controlled reads were aligned to genotype specific reference sequences (taxonomy ID's: AF297621 (A1), X02763 (A2), D00330 and AB073858 (B), AB033556 (C), X02496 (D), X75657 (E)), using the bwasw mapping option of the Burrows Wheeler Aligner (BWA) version 0.6.1-r104 with default settings. From the resulting sequencing alignment map (SAM) file coverage overviews per reference sequence position were generated disregarding nucleotides with a phred score below 30. Amino acid variation per position was determined disregarding codons containing >1 nucleotide with a phred < 20.

Multiple sequence alignment for Sequence Harmony

Aligned reads were collected for the 84 patients. Nucleotide sequences were translated and summarized as frequency counts per amino acid type per position per patient. A multiple sequence alignment between patients was performed based on consensus sequences per patient using the Praline multiple sequence analysis tool with global pre-profiling and default settings (Pirovano et al., 2008; Simossis et al., 2005). This multiple alignment of the consensus sequences was used to align frequency count tables between patients.

Comparison of viral sequences with Sequence Harmony

The Multi-Harmony (SHMR) webserver (www.ibi.vu.nl/programs/shmrwww) was used to calculate Sequence Harmony (SH) scores (Brandt et al., 2010; Pirovano et al., 2006), to analyze amino acid differences between the HBc sequences of the 84 CHB patients based

on the aligned frequency count tables. The SH algorithm is an entropy-based method, which measures the overlap in distribution of amino acid types between subgroups at positions within an alignment of related protein sequences divided into groups. An SH score of 0 indicates amino acid positions that are specific for one of the sequence groups, whereas an SH score of 1 indicates a complete overlap at this amino acid position between the two groups. For two subgroups (*A* and *B*), SH measures the overlap in distribution of amino acid types (*x*) at a certain position (*i*) in the sequence alignment using a formula based on relative entropy (mutual information) as follows:

$$\mathrm{SH}_i^{A/B} = \sum_x p_{\mathrm{i,x}}^A \log \frac{p_{\mathrm{i,x}}^A}{p_{\mathrm{i,x}}^A + \mathrm{p}_{\mathrm{i,x}}^B}$$

where $p_{\mathrm{i,x}}^A$ indicates the observed frequency in group A for amino acid type x at position i in the sequence, and $p_{\mathrm{i,x}}^B$ analogously for group B.

In our dataset, the observed frequencies *p* are not counted from the alignments of each group, as is done in the original method, but derived from the frequency per patient determined by next generation sequencing, as described above. Each patient will therefore have the same weight in (or impact on) the total score in each group independent of the number of sequence reads. Sequence coverage of the core protein is shown in **Appendix Figure 2**. To exclude spurious variations, e.g. arising from sequencing errors, we discounted all variants with a support of less than 50 reads. Most positions have a coverage of at least 5000 (**Appendix Figure 2**), which means that for a typical site, a minimum support of about 1% is effectively required. In addition, an empirical Z-score is calculated, reflecting the significance of the SH score obtained based on 100 random shuffling events of the sequences between the two groups. Cut-off scores for selecting positions significant for functional differences between the groups analysed, were set as SH < 0.960 and Z-score < -0.8. The high SH cut-off allows for detection of relatively small differences (>4%) against a background of overall high conservation. The Z-scores display the accuracy of a given result; the more negative the Z-score the less likely it is that the results arose by chance. For each HBV genotype, we compared HBeAg positive patients with HBeAg negative patients. Furthermore, we compared HBeAg positive and negative patients with a combined response (CR) to antiviral treatment to those who did not respond to treatment.

Amino acid variation in HBc

SH yielded a list of amino acid differences at selected positions in HBc between HBeAg positive and negative individuals, and a corresponding list for HBeAg negative individuals with and without treatment response. As per position these differences were relatively minor, we assessed the explanatory power of the SH selected positions combined. For this, we analyzed the amino acid variation at the SH selected positions in each group. For comparison of the amino acid variation at SH selected positions between HBeAg-negative and HBeAg-positive patients, the consensus amino acid sequence of HBeAg-positive patients was used as a reference to calculate the fraction of non-consensus sequence (% amino acid variation at that position). Similarly, for comparison of the amino acid variation at SH selected positions between responders and non-responders, the consensus amino acid sequence of non-responders was used as a reference. The average amino acid variation for the set of SH selected positions per patient was used to analyze the variation in HBc between the groups.

Structural analysis

Analysis of the protein three-dimensional structure informs us on the relative locations of the selected positions. Proximity in the structure indicates a possible shared functional role of the selected positions. We took the 3J2V X-ray cryo-EM structure of the HBc (Yu et al., 2013). This structure contains 4 copies of HBc in its assembled capsid conformation. The C-terminal region comprising residues 150 to 183 or 185 was not resolved. The sites selected by the SH analysis, i.e., the residues scoring below the SH and Z-score cut-offs, for HBeAg positive versus negative, and for HBeAg negative treatment response versus no response as described above, were mapped onto the three-dimensional structure. The mapped positions were clustered based closest distance between side-chain atoms (or the C_α for glycine); we used cut-off of 6.0Å, 7.0 Å and 8.0 Å. As we specifically analyze the residue side chains for the clustering, we also consider sequentially adjacent (consecutive) amino acids whose sidechains are within the cutoff as clusters. Mapped positions and clusters were visualized using PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

Statistical analyses

We tested for differences in amino acid variation between study groups using the Mann-Whitney U test, and correlations were analysed with Spearman's correlation coefficient (R) and multivariable linear regression using SPSS (version 23) and Graph Pad software (version 7.0).

RESULTS

Patient characteristics

Characteristics of CHB patients with 454 deep sequencing data available (n=89) are shown in **Table 1**. Patients were HBeAg positive (n=42) or HBeAg negative (n=47), and had HBV genotypes A (n=28), B (n=15), C (n=12), D (n=25), or E (n=9). Eighty-four patients completed treatment and follow-up and comprised the per-protocol population used for studying associations with treatment response. At week 72, 14/42 (35%) of HBeAg positive and 16/47 (36%) of HBeAg negative patients had combined response. HBsAg loss at 2 years of follow-up (week 144) was achieved in 5/42 (13%) of HBeAg positive and 8/47 (18%) of HBeAg negative patients (Jansen et al., 2014).

Sequence Harmony identifies viral variants in precore and HBc that correlate with HBeAg status

The SH method was used to identify amino acid differences in HBc between HBeAg positive and HBeAg negative patients with HBV genotypes A, B, C and D. HBV genotype E was excluded for this analysis since the number of HBeAg positive patients was too low (fewer than three patients). A low SH score indicates a position where the amino acid composition is different between the two study groups. In addition, an empirical Z-score is calculated based on 100-fold random permutation, to reflect the significance of the SH score (see Methods for details). The SH analysis yielded 54 sites with a SH score below the defined cut-off values (< 0.960), and a support of at least 50 read (corresponding to about 1% coverage, see **Appendix Figure 2**), indicating differences in amino acids present at these sites between the HBeAg positive and negative groups. These selected sites all have high (negative) Z-scores (< -0.8) **(Appendix Table 2)**, indicating that the minimum number of reads required of 50 was appropriate to exclude selection of spurious differences. As the minimum read depth of

our data for HBc was about 5000 (**Appendix Figure 2**), this allows the confident identification of minor variants of about 1%. Most of the observed amino acid differences between HBeAg positive and negative patients are present in minor viral populations. This is reflected in the relatively high median SH score for HBeAg status of 0.886 (IQR: 0.783 – 0.925) **(Figure 3)**, and that most of the differences in amino acid composition are present as minor viral variants (lowercase letters), while the major variants (uppercase letters) are often similar between both groups **(Appendix Table 2)**. Of the 54 positions that were identified by SH analysis, 23 positions were present in more than one HBV genotype. In the precore region, the amino acid composition was different between HBeAg positive and negative patients in the majority of HBV genotypes for residues at precore position 28 (genotypes A, B, D) and position 29 (genotypes A, C, D) **(Table 2)**. HBc region residues 49 (genotypes A, B, C, D), 77 (genotype A, B, C) and 130 (genotype A, B, D) had a different amino acid composition between HBeAg positive and negative patients in the majority of HBV genotypes **(Table 2)**.

Sequence Harmony identifies viral variants in HBc that correlate with therapy response

Most of the observed amino acid differences between combined responders and non-responders were present as minor viral variants, with a median SH score of 0.809 (IQR: 0.689 – 0.849) **(Appendix Figure 3)**, which indicates a relatively high similarity between the groups. First, we studied the amino acid differences in HBc in HBeAg positive patients with and without response to antiviral treatment. This analysis focused on HBV genotypes A and C, since the number of HBeAg positive responders was too low for HBV genotypes B, D and E. The SH analysis yielded 7 positions with differences in amino acids between HBeAg positive responders and non-responders (SH score < 0.960, Z-score < -0.8, and support ≥50 reads) **(Appendix Table 3)**. There was no overlap in the identified positions for HBV genotype A and C.

To identify amino acid differences in HBeAg negative responders versus non-responders, we have calculated SH scores for patients with HBV genotypes A, D and E. Patients with HBV genotypes B and C could not be assessed since the number of HBeAg negative responders was too low. The SH analysis yielded 22 sites with a different amino acid composition between HBeAg negative responders and non-responders (SH score < 0.960, Z-score < -0.8, and support ≥50 reads) **(Appendix Table 3)**. Of these 22 positions, only three positions were present in more than one HBV genotype (HBc residues 49, 80 and 114).

Structural mapping of the Sequence Harmony results reveals clustering of amino acids

To better understand the meaning of the SH results, we mapped the positions that differed significantly between study groups onto the three-dimensional structure of HBc. The 3J2V cryo-EM structure contains four chains in the unit cell, providing a detailed view of the HBc structure as well as most interactions that occur within the assembled icosahedral capsid. This allows us to analyse in detail the 24 SH selected residues per chain from the comparison between HBeAg negative and positive patients (HBV genotypes A and D); note that this is fewer than the 54 listed in **Appendix Table 2**, as there also positions for the other genotypes are included, and the available crystal structure does not include the pre-core domain, nor the C-terminal domain from position 147 onwards.

Figure 1A+B show these 24 selected sites, several of which show distinct clusters in the structure (sometimes by being adjacent in the sequence), mostly at the surface of the viral

capsid (shown in red and purple in **Figure 1C+D**) and some at the interfaces between the HBc capsid protein monomers (shown in red in **Figure 1A+B**). Around the four-helical bundle on the A-B interface, from top to bottom (**Figure 1A+B)**, we find the largest clusters. At a clustering cut-off of 6.0 Å (see Methods for details) for genotype A, selected positions 59 and 60 form a cluster near the center within each monomer chain, and positions 9, 49, 50, and 113-114 do likewise near the bottom (**Figure 1B+D**). In between those intra-chain clusters, positions 5 and 60 from chains A and B form a 5-residue inter-chain cluster (5A, 59B, 60B, 63B, 67B). At 7.0 Å, these merge into two big clusters (5A, 9A, 49A, 50A, 59B, 60B, 63B, 67B, 105B, and vice versa for B and A), also encompassing position 105. Other sites mainly cluster in twos or threes (at 6.0 Å): 130-131, and 21, 24 (at the left in chain A). One other inter-chain cluster for genotype A is noteworthy, albeit only at a cut-off of 8.0 Å: sites 40 and 45 between chains B and D (**Figure 1B+D)**. The selected positions for genotype D also add an inter-chain cluster at 8.0 Å: sites 12 and 40 between chains C and B.
For the comparison of HBeAg negative patients (HBV genotype A and D) with combined response versus non-response (shown in blue and purple in **Figure 1C+D**, and in blue in **Figure 1E+F**), observed differences were smaller, and the subsequently smaller set of selected residues also leads to fewer clusters, most of which are close together in the sequence and within the chain. For the genotype A selection, we only find a pair of residues clustering at a cut-off of 8.0 Å: 114 and 116. For genotype D at 6.0 Å, two clusters are obtained: 66 and 69, and positions 78, 79, and 80 of both chains A and D.
Clustering of SH selected residues from the comparison between HBeAg negative and positive patients for HBV genotype B and C are shown in **Appendix Table 4**.

<u>Higher amino acid variation in HBc in HBeAg negative patients compared to HBeAg positive patients</u>

Based on the positions revealed by the SH analysis **(Appendix Table 2)**, we calculated the amino acid variation (the fraction non-consensus sequence on selected positions) for HBeAg negative patients compared to the reference sequence of HBeAg positive patients. As a control, the amino acid variation for the HBeAg positive patients was calculated compared to its own reference sequence. The median amino acid variation of patients with various HBV genotypes was significantly higher for HBeAg negative patients versus HBeAg positive patients **(Figure 2A).** The presence of mutations in the BCP region, A1762T and G1764A, have been associated with HBeAg negative status (Alexopoulou et al., 1997; Buckwold et al., 1996). Mutations in BCP region of patients included in this analysis have previously been described (Jansen et al., 2017). Multivariable regression analysis showed that the association between the higher amino acid variation and an HBeAg negative status was independent of the presence of BCP mutations and HBV genotype (Table 3).

For comparison, we calculated the amino acid variation for the rest of HBc (entire HBc without positions included in the SH score). A significant correlation was found between the amino acid variation based on the SH selection and the rest of HBc (Spearman's R=0.3961, P=0.0005), independent of HBV genotype **(Figure 2B, Appendix Figure 4)**. The SH based amino acid variation for HBeAg status had a significant negative correlation with both pre-treatment HBV DNA level (Spearman's R= -0.7039, P<0.0001), HBsAg-level (Spearman's R= -0.5457, P<0.0001) **(Figure 2C+2D)**, and ALT level (Spearman's R= -0.2664, P=0.0218). The amino acid variation in HBc of patients with various HBV genotypes based on the SH results was not significantly different between HBeAg negative non-responders and responders **(Figure 3, Table 3)**.

DISCUSSION

This was the first study analysing amino acid differences of HBc in subgroups of CHB patients using deep sequencing and the SH algorithm. Differences in amino acid composition of HBc between HBeAg positive and negative patients were mostly found at the inner surface of the viral capsid and the intra-dimer interface, while most amino acid differences between HBeAg negative non-responders and responders were mainly located at the outer surface of the viral capsid. The amino acid variation as determined by the fraction non-consensus sequence on SH-selected positions, was higher in HBeAg negative patients, and was negatively correlated with HBV DNA levels and HBsAg levels.

The SH algorithm offers an way to quantify the overlap in amino acids between study groups (Brandt et al., 2010; Pirovano et al., 2008), which avoids personal biases that may arise when manually inspecting such alignments, and the a-priori bias on conserved positions that is applied by most other methods for detecting group-specific positions. The SH algorithm thus enables us to identify positions that may explain differences in disease outcome between patients. We observed a different amino acid composition between HBeAg positive and HBeAg negative CHB patients at PC residue 28, confirming a well-established point mutation from G to A at position 1896 of the PC region. This mutation converts the TGG codon for tryptophan to TAG, a translational stop codon, and thereby prevents the production of HBeAg which plays a role in suppressing immune response to HBV infection (Buckwold et al., 1996; Carman et al., 1989; Chen et al., 2004; Lok et al., 1994; Milich et al., 1990). Furthermore, we confirmed previous findings that sequence diversity in HBc is greater in HBeAg negative patients than HBeAg positive patients (Akarca and Lok, 1995; Chuang et al., 1993; Homs et al., 2014). (Akarca and Lok, 1995; Chuang et al., 1993; Homs et al., 2014). Although HBc and HBeAg are encoded by the same ORF, these proteins have a distinct structure and function (Zlotnick et al., 2013). While HBc proteins assemble into core particles encapsidising the viral RNA genome, the HBeAg protein contains a signal peptide and is secreted after processing. Mutations in the core gene ORF will result in amino acid changes affecting both HBc and HBeAg. In HBeAg negative patients, who have a viral variant expressing no or reduced amounts of HBeAg (Alexopoulou et al., 1997; Buckwold et al., 1996; Carman et al., 1989; Lok et al., 1994), amino acid changes in the core ORF will only affect the core protein. We observed that amino acid changes in HBc of HBeAg-negative patients were mostly found at the inner surface of the viral capsid. This suggest that the high selection pressure to conserve the amino acids at these positions is lost in HBeAg negative patients, most likely because these positions are essential to maintain HBeAg structure and function in HBeAg-positive patients. It is also possible that structures at the inner face of the core protein may be on the surface of the HBeAg (Zlotnick et al., 2013), therefore the HBeAg specific immune response may also play a role in the conservation of these residues in HBeAg positive patients. Furthermore, core mutations in HBeAg-negative patients could also lead to optimization of the HBc structure. We observed that differences in amino acid compositions associated with HBeAg status were found in both the assembly domain (in 39/149 residues, 26%), and in the CTD (in 9/34 residues, 26%) of HBc. Albeit the arginine-rich CTD is not necessary for the assembly of viral capsids (Lewellyn and Loeb, 2011; Nassal, 1992a), it has essential functions in the HBV replication cycle such as binding of nucleic acid, regulation of pgRNA packaging, rcDNA synthesis, and modulation of nuclear

import (Gazina et al., 2000; Kann and Gerlich, 1994; Lewellyn and Loeb, 2011; Liao and Ou, 1995; Nassal, 1992a; Petit and Pillot, 1985; Yeh and Ou, 1991). In line with previous studies, we found few variance in the arginine rich part of the CTD that is important for DNA binding, while we did observe minor amino acid variation in the arginine residues of the CTD involved in encapsidation of RNA (residues 150-154) in HBeAg negative patients (Akarca and Lok, 1995; Hatton et al., 1992). No amino acid variations were observed in the CTD serine (S) phosphorylation sites that are required for reverse transcription (S157, S164, and S170) (Gazina et al., 2000). In HBeAg negative patients, we did observe variation in an additional phosphoacceptor site with replacement of serine by threonine at residue 178 (Jung et al., 2014).

The structure of HBc is dominated by an α-helical hairpin (Wynne et al., 1999). The base of the protein is formed by a hydrophobic core which stabilizes the monomer fold, consisting of residues 6, 9, 15-16, 18-19, 23-24, 102-103, 110, 115, 118-119, 122, 125, and 140 (Wynne et al., 1999). Our study confirms that these hydrophobic residues are highly conserved (Wynne et al., 1999), with no evidence of coexistence of hydrophilic residues. However, we did observe amino acid differences in other residues at the base of the α-helix, such as leucine 60 (L60), that was previously shown to support intracellular capsid formation but abrogate virion secretion (Ponsel and Bruss, 2003). HBc dimers, required for the formation of the capsid, associate via their α-helical hairpins to form a four-helix bundle (Wynne et al., 1999). A disulphide bridge at the dimer-interface between the cysteine 61 (C61) residues stabilizes its association (Nassal, 1992b; Selzer et al., 2014). We did not observe differences in viral sequences between subgroups in the C61 position. Also, no relevant variations were observed in other residues which are known to be important for inter-subunit packaging such as arginine 127 (A127), proline 129 (P129), tyrosine 132 (Y132), and I139 (Wynne et al., 1999).

Previously, it was shown that replacement of phenylalanine/isoleucine (depending on HBV genotype) by leucine at residue 97 (FI/97L) resulted in secretion of virions containing immature (replication intermediate) single stranded (ss) DNA (Le Pogam and Shih, 2002; Yuan et al., 1999a). It was proposed that this immature virion secretion can be offset by another frequently occurring mutation; proline to threonine at position 130 (P130T) (Yuan and Shih, 2000). In our study, the dominant residue at position 97 was an isoleucine in HBV genotype B and C HBeAg positive patients, whereas this was an isoleucine, valine or leucine in HBeAg negative patients. Interestingly, we did not observe a viral population with a threonine at position 130 in HBeAg negative patients with genotypes B or C, suggesting that the FI/97L mutation is not always rescued by a second mutation in vivo. Mutations P5T and L60V, associated with inhibition of virion secretion possibly by interference with capsid envelopment (Le Pogam et al., 2000), were present as minor viral population in both HBeAg negative as HBeAg positive patients.

The presence of BCP and/or PC variants, as dominant or as minor population, was previously shown to limit the response to peg-IFN or NUC monotherapy in CHB patients (Bayliss et al., 2016; Erhardt et al., 2000; Jansen et al., 2017; Marrone et al., 2003; Sonneveld et al., 2012). In addition, our study demonstrates differences in several residue positions in HBc between HBeAg negative non-responders and combined responders to treatment with peg-IFN and NUCs. Interestingly, a substantial part of these residues were located at the outer surface of the viral capsid (residues 77-80), where the major immunodominant region of HBc has been mapped (Pumpens and Grens, 2001; Salfeld et al., 1989; Sallberg et al., 1991). This region lies at the tip of the spikes of the four-helix bundle

that binds to HBsAg, the inner surface of the envelop, which is essential for virus assembly (Bottcher et al., 1998; Dyson and Murray, 1995). HBV capsids are highly immunogenic, and are able to elicit B cell, T helper cell, and cytotoxic T cell responses. Therefore, amino acid substitutions in HBc may alter immune recognition and thereby affect immune clearance and treatment outcome (Akarca and Lok, 1995; Radecke et al., 2000). So far, a large number of HBV specific epitopes has been described (Desmond et al., 2008). The HBc 18-27 epitope (FLPSDFFPSV), evoking both CD4+ and CD8+ T cell responses, has been characterized extensively (Bertoletti et al., 1993; Bertoletti et al., 1997; Bertoni et al., 1997; Desmond et al., 2008; Hosono et al., 1995; Kefalakes et al., 2015; Lee et al., 1997; Nayersina et al., 1993; Rehermann et al., 1995a; Rehermann et al., 1995b). In this study, we observed several amino acid changes in this epitope, including one of the anchor residues, which were associated with both HBeAg status (V27T/I in HBeAg negative patients) and treatment response in HBeAg negative patients (V27A in responders). We were unable to analyse amino acid differences of HBc in patients with and without HBsAg loss, due to the limited number of patients with HBsAg loss in the study group.

HBeAg status has previously been associated with several markers of viral replication: HBeAg negative patients have lower HBsAg and HBV DNA levels in plasma and a higher viral diversity (Fattovich et al., 2008; Lim et al., 2007; Takkenberg et al., 2013; van de Klundert et al., 2012). In agreement, the amino acid variations at SH selected positions identified based on HBeAg status, were inversely correlated with HBV DNA and HBsAg levels. A greater variability in the HBc region in HBeAg negative patients than in HBeAg positive patients quasispecies was observed previously in a small number of CHB patients using deep sequencing techniques (Homs et al., 2014). It was postulated that an increased viral diversity may be associated with HBeAg seroconversion (Akarca and Lok, 1995; Homs et al., 2014), however in our analysis we did not take the time delay between HBeAg seroconversion and time of sequencing into account.

A strength of our study is that we have used samples of a well-characterized cohort of CHB patients including both HBeAg negative and positive patients. Furthermore, we have used deep sequencing techniques enabling to identify minority viral variants that may have been missed in earlier studies using population sequencing methods.

Based on this data, our analysis offers a better understanding of the possible drivers of HBc evolution in different subgroups of CHB patients. Amino acid variations in HBc may affect virion secretion, reverse transcription, or change important immunogenic regions of HBc. The two alternative fates of HBV cores, either envelopment followed by secretion, or transport to the host cell nucleus for release of rcDNA, require careful regulation of their stability (Cui et al., 2015). Finally, in the context of development of new antivirals for HBV, of which HBc allosteric modulators are promising candidates (Durantel and Zoulim, 2016; Klumpp et al., 2015; Yang et al., 2016), it is essential to gain better insight in the different properties of HBc to better understand the underlying mechanism of these HBc targeting antivirals.

REFERENCES

Akarca, U.S., Lok, A.S., 1995. Naturally occurring hepatitis B virus core gene mutations. Hepatology 22, 50-60.

Alexopoulou, A., Karayiannis, P., Hadziyannis, S.J., Aiba, N., Thomas, H.C., 1997. Emergence and selection of HBV variants in an anti-HBe positive patient persistently infected with quasi-species. J Hepatol 26, 748-753.

Bayliss, J., Yuen, L., Rosenberg, G., Wong, D., Littlejohn, M., Jackson, K., Gaggar, A., Kitrinos, K.M., Subramanian, G.M., Marcellin, P., Buti, M., Janssen, H.L., Gane, E., Sozzi, V., Colledge, D., Hammond, R., Edwards, R., Locarnini, S., Thompson, A., Revill, P.A., 2016. Deep sequencing shows that HBV basal core promoter and precore variants reduce the likelihood of HBsAg loss following tenofovir disoproxil fumarate therapy in HBeAg-positive chronic hepatitis B. Gut.

Bertoletti, A., Chisari, F.V., Penna, A., Guilhot, S., Galati, L., Missale, G., Fowler, P., Schlicht, H.J., Vitiello, A., Chesnut, R.C., et al., 1993. Definition of a minimal optimal cytotoxic T-cell epitope within the hepatitis B virus nucleocapsid protein. J Virol 67, 2376-2380.

Bertoletti, A., Southwood, S., Chesnut, R., Sette, A., Falco, M., Ferrara, G.B., Penna, A., Boni, C., Fiaccadori, F., Ferrari, C., 1997. Molecular features of the hepatitis B virus nucleocapsid T-cell epitope 18-27: interaction with HLA and T-cell receptor. Hepatology 26, 1027-1034.

Bertoni, R., Sidney, J., Fowler, P., Chesnut, R.W., Chisari, F.V., Sette, A., 1997. Human histocompatibility leukocyte antigen-binding supermotifs predict broadly cross-reactive cytotoxic T lymphocyte responses in patients with acute hepatitis. J Clin Invest 100, 503-513.

Birnbaum, F., Nassal, M., 1990. Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. J Virol 64, 3319-3330.

Bottcher, B., Tsuji, N., Takahashi, H., Dyson, M.R., Zhao, S., Crowther, R.A., Murray, K., 1998. Peptides that block hepatitis B virus assembly: analysis by cryomicroscopy, mutagenesis and transfection. EMBO J 17, 6839-6845.

Brandt, B.W., Feenstra, K.A., Heringa, J., 2010. Multi-Harmony: detecting functional specificity from sequence alignment. Nucleic Acids Res 38, W35-40.

Buckwold, V.E., Xu, Z., Chen, M., Yen, T.S., Ou, J.H., 1996. Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. J Virol 70, 5845-5851.

Carman, W.F., Jacyna, M.R., Hadziyannis, S., Karayiannis, P., McGarvey, M.J., Makris, A., Thomas, H.C., 1989. Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. Lancet 2, 588-591.

Chen, M.T., Billaud, J.N., Sallberg, M., Guidotti, L.G., Chisari, F.V., Jones, J., Hughes, J., Milich, D.R., 2004. A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. Proc Natl Acad Sci U S A 101, 14913-14918.

Chevaliez, S., Hezode, C., Bahrami, S., Grare, M., Pawlotsky, J.M., 2013. Long-term hepatitis B surface antigen (HBsAg) kinetics during nucleoside/nucleotide analogue therapy: finite treatment duration unlikely. J Hepatol 58, 676-683.

Chuang, W.L., Omata, M., Ehata, T., Yokosuka, O., Ito, Y., Imazeki, F., Lu, S.N., Chang, W.Y., Ohto, M., 1993. Precore mutations and core clustering mutations in chronic hepatitis B virus infection. Gastroenterology 104, 263-271.

Cui, X., Luckenbaugh, L., Bruss, V., Hu, J., 2015. Alteration of Mature Nucleocapsid and Enhancement of Covalently Closed Circular DNA Formation by Hepatitis B Virus Core Mutants Defective in Complete-Virion Formation. J Virol 89, 10064-10072.

Desmond, C.P., Bartholomeusz, A., Gaudieri, S., Revill, P.A., Lewin, S.R., 2008. A systematic review of T-cell epitopes in hepatitis B virus: identification, genotypic variation and relevance to antiviral therapeutics. Antivir Ther 13, 161-175.

Dienstag, J.L., 2008. Hepatitis B virus infection. N Engl J Med 359, 1486-1500.

Durantel, D., Zoulim, F., 2016. New antiviral targets for innovative treatment concepts for hepatitis B virus and hepatitis delta virus. J Hepatol 64, S117-S131.

Dyson, M.R., Murray, K., 1995. Selection of peptide inhibitors of interactions involved in complex protein assemblies: association of the core and surface antigens of hepatitis B virus. Proc Natl Acad Sci U S A 92, 2194-2198.

Erhardt, A., Reineke, U., Blondin, D., Gerlich, W.H., Adams, O., Heintges, T., Niederau, C., Haussinger, D., 2000. Mutations of the core promoter and response to interferon treatment in chronic replicative hepatitis B. Hepatology 31, 716-725.

Fattovich, G., Bortolotti, F., Donato, F., 2008. Natural history of chronic hepatitis B: special emphasis on disease progression and prognostic factors. J Hepatol 48, 335-352.

Gazina, E.V., Fielding, J.E., Lin, B., Anderson, D.A., 2000. Core protein phosphorylation modulates pregenomic RNA encapsidation to different extents in human and duck hepatitis B viruses. J Virol 74, 4721-4728.

Guo, Y.H., Li, Y.N., Zhao, J.R., Zhang, J., Yan, Z., 2011. HBc binds to the CpG islands of HBV cccDNA and promotes an epigenetic permissive state. Epigenetics 6, 720-726.

Hatton, T., Zhou, S., Standring, D.N., 1992. RNA- and DNA-binding activities in hepatitis B virus capsid protein: a model for their roles in viral replication. J Virol 66, 5232-5241.

Homs, M., Caballero, A., Gregori, J., Tabernero, D., Quer, J., Nieto, L., Esteban, R., Buti, M., Rodriguez-Frias, F., 2014. Clinical application of estimating hepatitis B virus quasispecies complexity by massive sequencing: correlation between natural evolution and on-treatment evolution. PLoS One 9, e112306.

Hosono, S., Tai, P.C., Wang, W., Ambrose, M., Hwang, D.G., Yuan, T.T., Peng, B.H., Yang, C.S., Lee, C.S., Shih, C., 1995. Core antigen mutations of human hepatitis B virus in hepatomas accumulate in MHC class II-restricted T cell epitopes. Virology 212, 151-162.

Jansen, L., de Niet, A., Stelma, F., van Iperen, E.P., van Dort, K.A., Plat-Sinnige, M.J., Takkenberg, R.B., Chin, D.J., Zwinderman, A.H., Lopatin, U., Kootstra, N.A., Reesink, H.W., 2014. HBsAg loss in patients treated with peginterferon alfa-2a and adefovir is associated with SLC16A9 gene variation and lower plasma carnitine levels. J Hepatol 61, 730-737.

Jansen, L., Welkers, M.R.A., van Dort, K.A., Takkenberg, R.B., Lopatin, U., Zaaijer, H.L., de Jong, M.D., Reesink, H.W., Kootstra, N.A., 2017. Viral minority variants in the core promoter and precore region identified by deep sequencing are associated with response to peginterferon and adefovir in HBeAg negative chronic hepatitis B patients. Antiviral Res 145, 87-95.

Janssen, H.L., van Zonneveld, M., Senturk, H., Zeuzem, S., Akarca, U.S., Cakaloglu, Y., Simon, C., So, T.M., Gerken, G., de Man, R.A., Niesters, H.G., Zondervan, P., Hansen, B., Schalm, S.W., Group, H.B.V.S., Rotterdam Foundation for Liver, R., 2005. Pegylated interferon alfa-2b alone or in combination with lamivudine for HBeAg-positive chronic hepatitis B: a randomised trial. Lancet 365, 123-129.

Jung, J., Hwang, S.G., Chwae, Y.J., Park, S., Shin, H.J., Kim, K., 2014. Phosphoacceptors threonine 162 and serines 170 and 178 within the carboxyl-terminal RRRS/T motif of the hepatitis B virus core protein make multiple contributions to hepatitis B virus replication. J Virol 88, 8754-8767.

Kann, M., Gerlich, W.H., 1994. Effect of core protein phosphorylation by protein kinase C on encapsidation of RNA within core particles of hepatitis B virus. J Virol 68, 7993-8000.

Kefalakes, H., Budeus, B., Walker, A., Jochum, C., Hilgard, G., Heinold, A., Heinemann, F.M., Gerken, G., Hoffmann, D., Timm, J., 2015. Adaptation of the hepatitis B virus core protein to CD8(+) T-cell selection pressure. Hepatology 62, 47-56.

Klumpp, K., Lam, A.M., Lukacs, C., Vogel, R., Ren, S., Espiritu, C., Baydo, R., Atkins, K., Abendroth, J., Liao, G., Efimov, A., Hartman, G., Flores, O.A., 2015. High-resolution crystal structure of a hepatitis B virus replication inhibitor bound to the viral core protein. Proc Natl Acad Sci U S A 112, 15196-15201.

Koschel, M., Oed, D., Gerelsaikhan, T., Thomssen, R., Bruss, V., 2000. Hepatitis B virus core gene mutations which block nucleocapsid envelopment. J Virol 74, 1-7.

Le Pogam, S., Shih, C., 2002. Influence of a putative intermolecular interaction between core and the pre-S1 domain of the large envelope protein on hepatitis B virus secretion. J Virol 76, 6510-6517.

Le Pogam, S., Yuan, T.T., Sahu, G.K., Chatterjee, S., Shih, C., 2000. Low-level secretion of human hepatitis B virus virions caused by two independent, naturally occurring mutations (P5T and L60V) in the capsid protein. J Virol 74, 9099-9105.

Lee, H.G., Lim, J.S., Lee, K.Y., Choi, Y.K., Choe, I.S., Chung, T.W., Kim, K., 1997. Peptide-specific CTL induction in HBV-seropositive PBMC by stimulation with peptides in vitro: novel epitopes identified from chronic carriers. Virus Res 50, 185-194.

Lewellyn, E.B., Loeb, D.D., 2011. The arginine clusters of the carboxy-terminal domain of the core protein of hepatitis B virus make pleiotropic contributions to genome replication. J Virol 85, 1298-1309.

Li, H.C., Huang, E.Y., Su, P.Y., Wu, S.Y., Yang, C.C., Lin, Y.S., Chang, W.C., Shih, C., 2010. Nuclear export and import of human hepatitis B virus capsid protein and particles. PLoS Pathog 6, e1001162.

Liao, W., Ou, J.H., 1995. Phosphorylation and nuclear localization of the hepatitis B virus core protein: significance of serine in the three repeated SPRRR motifs. J Virol 69, 1025-1029.

Lim, S.G., Cheng, Y., Guindon, S., Seet, B.L., Lee, L.Y., Hu, P., Wasser, S., Peter, F.J., Tan, T., Goode, M., Rodrigo, A.G., 2007. Viral quasi-species evolution during hepatitis Be antigen seroconversion. Gastroenterology 133, 951-958.

Lok, A.S., Akarca, U., Greene, S., 1994. Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. Proc Natl Acad Sci U S A 91, 4077-4081.

Marcellin, P., Lau, G.K., Bonino, F., Farci, P., Hadziyannis, S., Jin, R., Lu, Z.M., Piratvisuth, T., Germanidis, G., Yurdaydin, C., Diago, M., Gurel, S., Lai, M.Y., Button, P., Pluck, N., Peginterferon Alfa-2a, H.-N.C.H.B.S.G., 2004. Peginterferon alfa-2a alone, lamivudine alone, and the two in combination in patients with HBeAg-negative chronic hepatitis B. N Engl J Med 351, 1206-1217.

Marrone, A., Zampino, R., Luongo, G., Utili, R., Karayiannis, P., Ruggiero, G., 2003. Low HBeAg serum levels correlate with the presence of the double A1762T/G1764A core promoter mutation and a positive response to interferon in patients with chronic hepatitis B virus infection. Intervirology 46, 222-226.

Milich, D.R., Jones, J.E., Hughes, J.L., Price, J., Raney, A.K., McLachlan, A., 1990. Is a function of the secreted hepatitis B e antigen to induce immunologic tolerance in utero? Proc Natl Acad Sci U S A 87, 6599-6603.

Nassal, M., 1992a. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. J Virol 66, 4107-4116.

Nassal, M., 1992b. Conserved cysteines of the hepatitis B virus core protein are not required for assembly of replication-competent core particles nor for their envelopment. Virology 190, 499-505.

Nayersina, R., Fowler, P., Guilhot, S., Missale, G., Cerny, A., Schlicht, H.J., Vitiello, A., Chesnut, R., Person, J.L., Redeker, A.G., Chisari, F.V., 1993. HLA A2 restricted cytotoxic T lymphocyte responses to multiple hepatitis B surface antigen epitopes during hepatitis B virus infection. J Immunol 150, 4659-4671.

Petit, M.A., Pillot, J., 1985. HBc and HBe antigenicity and DNA-binding activity of major core protein P22 in hepatitis B virus core particles isolated from the cytoplasm of human liver cells. J Virol 53, 543-551.

Pirovano, W., Feenstra, K.A., Heringa, J., 2006. Sequence comparison by sequence harmony identifies subtype-specific functional sites. Nucleic Acids Res 34, 6540-6548.

Pirovano, W., Feenstra, K.A., Heringa, J., 2008. PRALINETM: a strategy for improved multiple alignment of transmembrane proteins. Bioinformatics 24, 492-497.

Ponsel, D., Bruss, V., 2003. Mapping of amino acid side chains on the surface of hepatitis B virus capsids required for envelopment and virion formation. J Virol 77, 416-422.

Pumpens, P., Grens, E., 2001. HBV core particles as a carrier for B cell/T cell epitopes. Intervirology 44, 98-114.

Radecke, K., Protzer, U., Trippler, M., Meyer Zum Buschenfelde, K.H., Gerken, G., 2000. Selection of hepatitis B virus variants with aminoacid substitutions inside the core antigen during interferon-alpha therapy. J Med Virol 62, 479-486.

Rehermann, B., Fowler, P., Sidney, J., Person, J., Redeker, A., Brown, M., Moss, B., Sette, A., Chisari, F.V., 1995a. The cytotoxic T lymphocyte response to multiple hepatitis B virus polymerase epitopes during and after acute viral hepatitis. J Exp Med 181, 1047-1058.
Rehermann, B., Pasquinelli, C., Mosier, S.M., Chisari, F.V., 1995b. Hepatitis B virus (HBV) sequence variation of cytotoxic T lymphocyte epitopes is not common in patients with chronic HBV infection. J Clin Invest 96, 1527-1534.
Salfeld, J., Pfaff, E., Noah, M., Schaller, H., 1989. Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. J Virol 63, 798-808.
Sallberg, M., Ruden, U., Magnius, L.O., Harthus, H.P., Noah, M., Wahren, B., 1991. Characterisation of a linear binding site for a monoclonal antibody to hepatitis B core antigen. J Med Virol 33, 248-252.
Selzer, L., Katen, S.P., Zlotnick, A., 2014. The hepatitis B virus core protein intradimer interface modulates capsid assembly and stability. Biochemistry 53, 5496-5504.
Simossis, V.A., Kleinjung, J., Heringa, J., 2005. Homology-extended sequence alignment. Nucleic Acids Res 33, 816-824.
Sonneveld, M.J., Rijckborst, V., Zeuzem, S., Heathcote, E.J., Simon, K., Senturk, H., Pas, S.D., Hansen, B.E., Janssen, H.L., 2012. Presence of precore and core promoter mutants limits the probability of response to peginterferon in hepatitis B e antigen-positive chronic hepatitis B. Hepatology 56, 67-75.
Summers, J., Mason, W.S., 1982. Replication of the genome of a hepatitis B--like virus by reverse transcription of an RNA intermediate. Cell 29, 403-415.
Takkenberg, R.B., Jansen, L., de Niet, A., Zaaijer, H.L., Weegink, C.J., Terpstra, V., Dijkgraaf, M.G., Molenkamp, R., Jansen, P.L., Koot, M., Rijckborst, V., Janssen, H.L., Beld, M.G., Reesink, H.W., 2013. Baseline hepatitis B surface antigen (HBsAg) as predictor of sustained HBsAg loss in chronic hepatitis B patients treated with pegylated interferon-alpha2a and adefovir. Antivir Ther 18, 895-904.
Tong, S., Revill, P., 2016. Overview of hepatitis B viral replication and genetic variability. J Hepatol 64, S4-S16.
van de Klundert, M.A., Cremer, J., Kootstra, N.A., Boot, H.J., Zaaijer, H.L., 2012. Comparison of the hepatitis B virus core, surface and polymerase gene substitution rates in chronically infected patients. J Viral Hepat 19, e34-40.
Watts, N.R., Conway, J.F., Cheng, N., Stahl, S.J., Belnap, D.M., Steven, A.C., Wingfield, P.T., 2002. The morphogenic linker peptide of HBV capsid protein forms a wmobile array on the interior surface. EMBO J 21, 876-884.
Wynne, S.A., Crowther, R.A., Leslie, A.G., 1999. The crystal structure of the human hepatitis B virus capsid. Mol Cell 3, 771-780.
Yang, L., Wang, Y.J., Chen, H.J., Shi, L.P., Tong, X.K., Zhang, Y.M., Wang, G.F., Wang, W.L., Feng, C.L., He, P.L., Xu, Y.B., Lu, M.J., Tang, W., Nan, F.J., Zuo, J.P., 2016. Effect of a hepatitis B virus inhibitor, NZ-4, on capsid formation. Antiviral Res 125, 25-33.
Yeh, C.T., Ou, J.H., 1991. Phosphorylation of hepatitis B virus precore and core proteins. J Virol 65, 2327-2331.
Yu, X., Jin, L., Jih, J., Shih, C., Zhou, Z.H., 2013. 3.5A cryoEM structure of hepatitis B virus core assembled from full-length core protein. PLoS One 8, e69729.
Yuan, T.T., Sahu, G.K., Whitehead, W.E., Greenberg, R., Shih, C., 1999a. The mechanism of an immature secretion phenotype of a highly frequent naturally occurring missense mutation at codon 97 of human hepatitis B virus core antigen. J Virol 73, 5731-5740.
Yuan, T.T., Shih, C., 2000. A frequent, naturally occurring mutation (P130T) of human hepatitis B virus core antigen is compensatory for immature secretion phenotype of another frequent variant (I97L). J Virol 74, 4929-4932.
Yuan, T.T., Tai, P.C., Shih, C., 1999b. Subtype-independent immature secretion and subtype-dependent replication deficiency of a highly frequent, naturally occurring mutation of human hepatitis B virus core antigen. J Virol 73, 10122-10128.
Zlotnick, A., Cheng, N., Conway, J.F., Booy, F.P., Steven, A.C., Stahl, S.J., Wingfield, P.T., 1996. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. Biochemistry 35, 7412-7421.
Zlotnick, A., Cheng, N., Stahl, S.J., Conway, J.F., Steven, A.C., Wingfield, P.T., 1997. Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein: implications for morphogenesis and organization of encapsidated RNA. Proc Natl Acad Sci U S A 94, 9556-9561.

Table 1: Baseline characteristics. Variables are shown for patients included in the present study with 454 DS data available (N=89) according to HBeAg status

	Baseline HBeAg status	
	HBeAg positive (N=42)	**HBeAg negative (N=47)**
Demographics		
Female, n (%)	8 (19)	15 (32)
Mean age, years (SD)	35.76 (9.45)	43.07 (9.84)
Ethnicity		
Caucasian, n (%)	16 (38)	12 (26)
Asian, n (%)	19 (45)	14 (30)
African, n (%)	7 (17)	21 (45)
IFN naïve, n (%)	33 (79)	33 (70)
Laboratory characteristics		
Median ALT, U/L (IQR)	97 (50-215)	63 (47-118)
HBV genotype		
A, n (%)	17 (40)	11 (23)
B, n (%)	8 (19)	7 (15)
C, n (%)	7 (17)	5 (11)
D, n (%)	8 (19)	17 (36)
E, n (%)	2 (5)	7 (15)
Mean HBV DNA, log10 IU/mL (SD)	8.02 (1.23)	5.56 (1.08)
Mean HBsAg, log10 IU/mL (SD)	4.29 (0.74)	3.33 (0.67)
Mean HBeAg, log10 IU/mL (SD)	2.61 (1.03)	n.a.
Treatment response [a]		
Combined response at week 72, n (%)	14 (35)	16 (36)
HBsAg loss at week 144, n (%)	5 (13)	8 (18)

n.a.: not applicable; a. Therapy response was determined in the per-protocol population (n=84) excluding 2 HBeAg-positive and 3 HBeAg-negative patients who did not complete the study treatment and follow-up.

Table 2: Amino acid differences between HBeAg positive and negative patients occurring in majority of HBV genotypes

Residue	HBV GT	SH-score	Z-score	HBeAg pos [a]	HBeAg neg [a]
Precore [b]					
28	A	0.869	-1.960	W(*)	W*
"	B	0.548	-10.593	W*	*w
"	D	0.241	-15.815	W*	*(w)
29	A	0.861	-2.382	G(d)	Gd
"	C	0.747	-1.400	Gs(d)	DGns
"	D	0.785	-7.915	Gd	GD
Core [c]					
49	A	0.934	-1.585	S	St
"	B	0.743	-2.045	S	TS
"	C	0.689	-2.027	S	ST
"	D	0.896	-1.652	S	St
77	A	0.520	-6.940	Eq	Qde(n)
"	B	0.874	-1.387	E	Ed
"	C	0.451	-3.562	E	DEQ
130	A	0.941	-1.265	P(l)	Pq
"	B	0.911	-1.878	P	Pa
"	D	0.835	-0.894	P(s)	Plq(s)

Sequence Harmony (SH) score < 0.960 and Z-score < - 0.8;
a. Summary of amino acid type frequencies in the indicated groups, sorted starting from the most frequently occurring type, lower case letters are used for a frequency less than half of the most occurring, and between brackets for a frequency of less than 1%; *: stopcodon;
b. Numbering from start of precore in genotype A reference sequence;
c. Numbering from start of core in genotype A reference sequence.

Table 3. Multivariable regression analysis of amino acid variation in subgroups of chronic hepatitis B patients

	Variable	B[a]	p-value	95% CI
HBeAg status				
	Amino acid variation at SH-selected positions	3,56	0,001	1,51 – 5,62
	BCP-mutations[b]	0,192	0,001	0,084 – 0,299
	HBV-genotype	0,069	ns	-0,009 – 0,147
Treatment response in HBeAg-negative				
	Amino acid variation at SH-selected positions	0,297	ns	-3,3 – 3,9
	HBV-genotype	-0,027	ns	-0,15 – 0,09

a. Unstandardized coefficient (B) indicates the difference in the outcome (HBeAg status or treatment response in HBeAg-negative) with respect to the variables. b. BCP-mutations were previously described (Jansen et al., 2017).

Figure 1

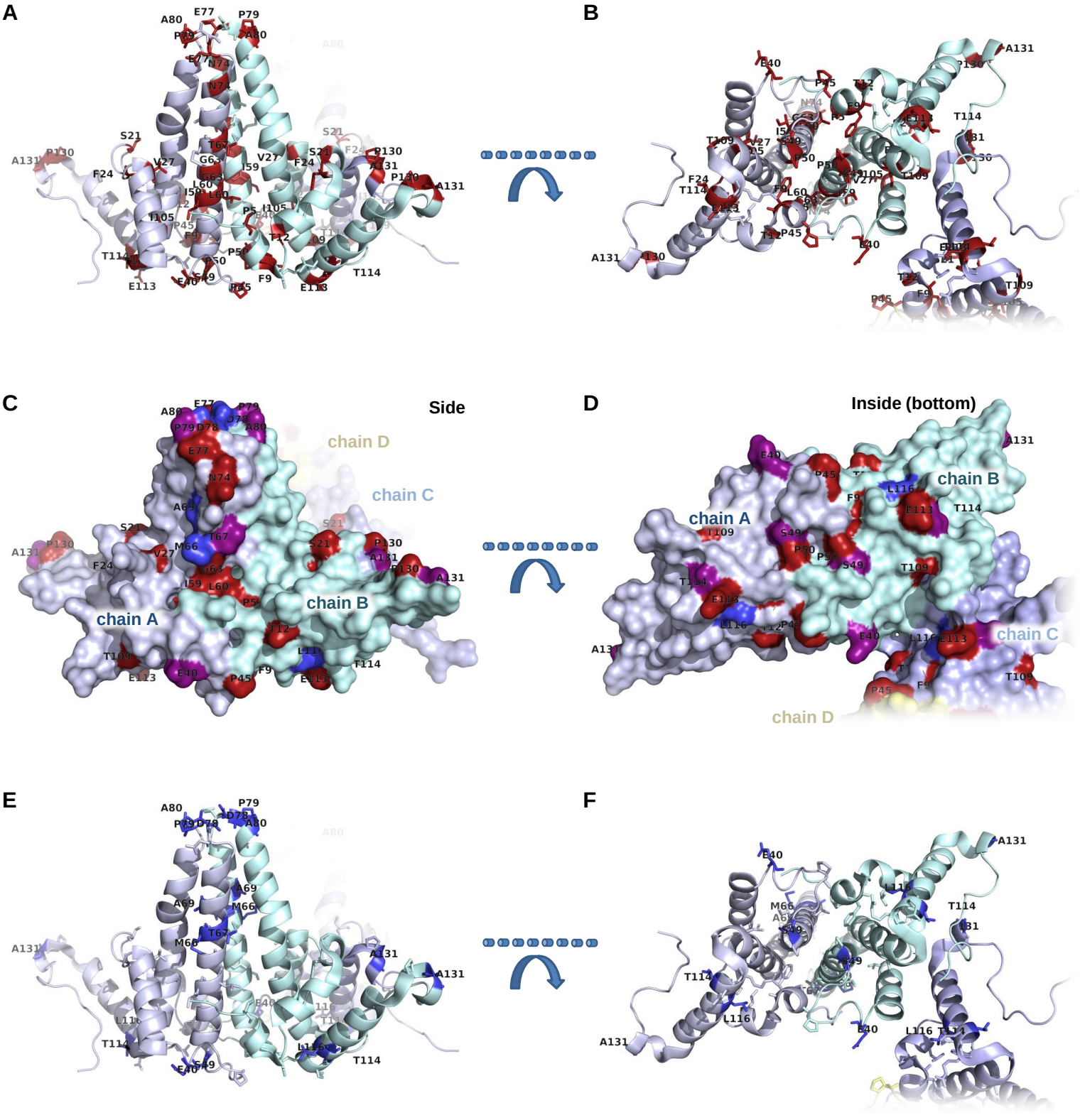
A
B
C
Side
chain D
chain C
chain B
chain A
D
Inside (bottom)
chain B
chain A
chain C
chain D
E
F

Fig 1. Structural mapping of SH selected sites, onto the four chains contained in the 3J2V HBV core protein crystal structure. The structure is arranged as two dimers: a dimer of A with B, and of C with D, which are connected by interactions between chains B and C. Both dimers (A-B and C-D) are two-fold rotation symmetric around the 'top-bottom' axis, which is perpendicular the capsid surface. We show the protein in two representations: cartoon (panel A+B+D+E) and surface (panel C+D); and in two orientations: side, i.e. tangential to the capsid surface, with the A-B dimer two-fold rotational symmetry axis pointing up (left-hand column), and inside, i.e. looking out from the capsid interior outwards, with the A-B axis pointing away from the viewer (right-hand column). Because of the symmetry, the visible side of chain A corresponds to opposite (back) side of chain B. And because both A-B and C-D dimers are largely identical, we show mainly chains A and B, and the interactions from B to C. In the cartoon representation, selected sites also drawn as sticks. Selected sites labelled by their single letter amino acid and reference position. Chains are labelled and indicated by different pastel shades in the surface representation. (**1A+1B)** 24 selected sites for HBeAg positive versus negative patients (HBV genotype A and D) are shown in red; **(1C+1D)** Both selections (HBeAg status and treatment response) are shown; sites selected by both analyses are indicated in purple; (**1E+1F)** 11 selected sites for HBeAg negative responders versus non-responders (HBV genotype A and D) are shown in blue.

Figure 2

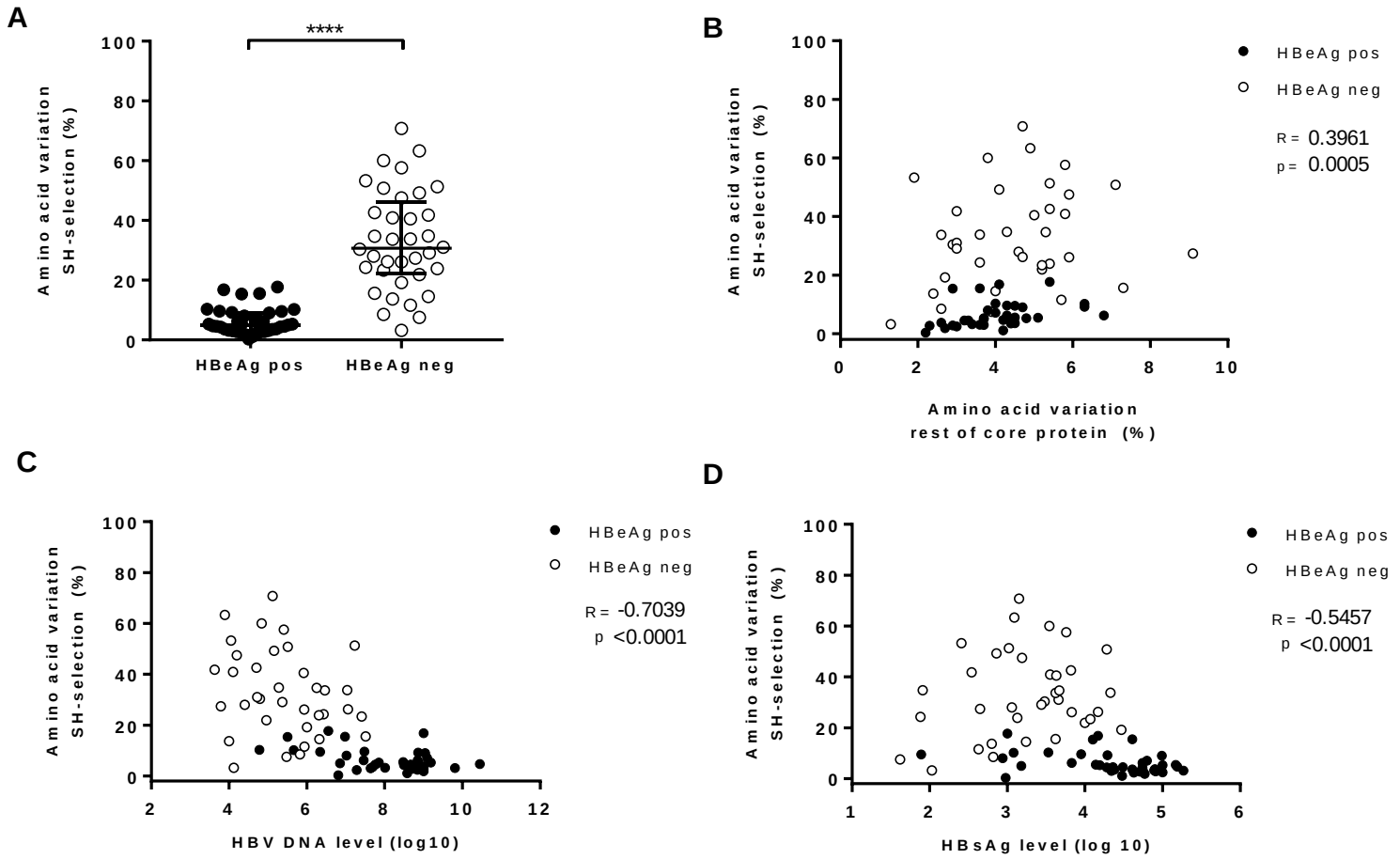

Fig 2. (2A) Amino acid variation based on Sequence Harmony (SH) selection is higher in HBeAg negative patients (white dots) compared to HBeAg positive patients (black dots). Median and interquartile range (IQR) are shown. Mann-Whitney U test was used to compare study groups, **** $p < 0.0001$. **(2B)** HBeAg negative patients have a higher amino acid variation in both SH selection and rest of HBV core protein compared to HBeAg positive patients. There was a significant correlation between SH based amino acid variation and rest of HBV core protein in HBeAg negative patients. Significant correlation between SH based amino acid variation and HBV DNA level **(2C)** and HBsAg level **(2D)**. Spearman's Rank test for correlation was used.

Figure 3

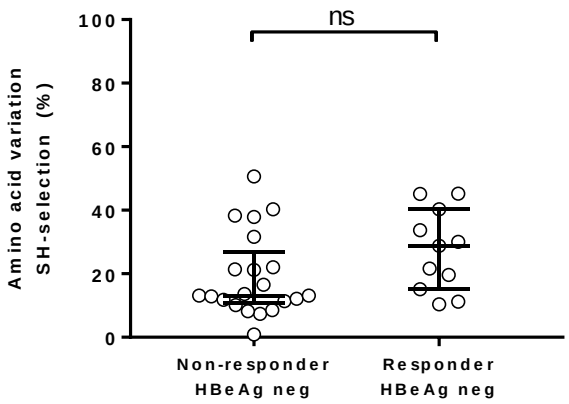

Fig 3. No significant difference in median Sequence Harmony based amino acid variation between HBeAg negative non-responders and responders to antiviral therapy. Median and interquartile range (IQR) are shown. Mann-Whitney U test was used to compare study groups.